

# School of Informatics



## Informatics Project Proposal Data Education for Children

**B137119**  
**April 2019**

### **Abstract**

In this project proposal, we introduce the motivation, background, methodology, and expected outcomes of our project - data education for children. We determine detailed milestones that can be used to measure its progress, and divide the entire project into dependent work packages. We also specify the specific completion time of each task.

Date: Thursday 11<sup>th</sup> April, 2019

**Tutor:** Pablo Leon Villagra  
**Supervisor:** Judy Robertson

# Contents

<b>1</b>	<b>Motivation</b>	<b>1</b>
1.1	Problem Statement . . . . .	2
1.2	Research Objectives . . . . .	2
1.3	Timeliness and Novelty . . . . .	2
1.4	Significance . . . . .	3
1.5	Feasibility . . . . .	3
1.6	Beneficiaries . . . . .	3
<b>2</b>	<b>Background</b>	<b>4</b>
<b>3</b>	<b>Programme and Methodology</b>	<b>4</b>
<b>4</b>	<b>Evaluation</b>	<b>5</b>
<b>5</b>	<b>Expected Outcomes</b>	<b>5</b>
<b>6</b>	<b>Research Plan, Milestones and Deliverables</b>	<b>6</b>

## 1 Motivation

Nowadays, most countries are making great efforts to the development of information technology. [1] analyzed new data on IT investment and productivity for 45 countries from 1994 to 2007, and compare the results with earlier research. It is found that upper-income developing countries have achieved positive and significant productivity gains from IT investment in the more recent period as they have increased their IT capital stocks and gained experience with the use of IT. As a sub-discipline of informatics, data science is developing rapidly in recent years. Data science can increase efficiency and productivity. Using data science can create organizational improvements of companies, public institutions, NGOs and even social movements [2]. Children are the future of a country, which determines the future development of the country to a certain extent. The general subject area of our project is children's informatics education. One understanding data science can process and interpret data as it pertains to answering real-world questions [3]. Hence, we hope that children can start learning data science in the early period. The project "Data Education for Children" came into being. We are committed to develop a series of teaching materials to empower children aged 10-12 to understand data science, show them cutting-edge data visualisation methods, and teach them basic data visualisation methods.

The project proposal is organized as follows: section 1 presents the motivation of our project; section 2 introduces the background of data education for children, including a knowledge and understanding of past and current work in children's informatics education; section 3 interprets the programme and methodology of accomplishing the project; section 4 presents two evaluation methods we will use for testing the quality of teaching materials ; section 5 discusses the expected outcomes; section 6 concludes the research plan, milestones and deliverables.

## 1.1 Problem Statement

Currently, most universities have data science courses designed for undergraduates (students over 18 years old). Scottish government set up the Data Lab MSc, which is a collaborative effort between seven Scottish universities with the aim of developing the data science talent and skills required by industry in Scotland [4]. Most primary and secondary schools do not offer data science courses for students of the corresponding age. At present, there is no widely acclaimed teaching material for children about data science. This has led most students unable to have a complete understanding of data science before the age of 18, being the gap that our project is committed to fill in.

## 1.2 Research Objectives

Our overall aim is to develop a set of interactive teaching materials which convey key techniques in data science in an interesting and approachable way. To achieve this aim, the overall aim is divided into three measurable objectives. The first objective is the data set used in the teaching material. The data set needs to be interesting enough (enough to engage children, e.g., pet-related data sets), and let children learn something from it. The second objective is the teaching method. We will analyze teaching methods related to programming education and visualisation literacy education to explore teaching methods suitable for data education. The third objective is the teaching material. The teaching material needs to be interesting, practical, and interactive.

Our project is an exploratory qualitative work. The evaluation of the teaching material is mainly based on the feedback from the users (children) of the teaching material. The specific evaluation methods will be introduced in section 4.

Our project aims to give children a certain concept of data science, develop their sensitivity to numbers, and cultivate their interest in data science. We may use the existing good methods of programming education, yet our teaching material does not involve programming.

## 1.3 Timeliness and Novelty

At present, many elementary schools have opened courses in information education (computer, programming, etc.), and children are also offered sophisticated teaching materials. But the core knowledge of data science - data cleaning, data visualization, data analysis, and data applying, is not included in current informatics courses. There is currently no authoritative teaching material suitable for children to learn data science. So, at present, developing the teaching material of the data science for the children is still a timely project.

Children are in the midst of looking for new things. Data science is an extremely novel thing for children. If our teaching material provide them with interesting data sets and interesting visualizations, the teaching material will have a strong novelty. Moreover, the same data set has various visualization methods. In the process of contacting these images, children can also learn the ability of divergent thinking. They can learn the basic methods of data science while touching novel things.

## 1.4 Significance

At present, some teaching materials and courses intended for children about computer science are available online. (e.g., course “STEM CHALLENGE: WRITE COMPUTER CODE” at [thestemlaboratory.com](http://thestemlaboratory.com) offers guidelines and packages, explaining what “binary computer code” means and how computers, CDs, DVDs and apps use it to communicate) Whereas currently there is no authoritative teaching material suitable for children to learn data science. Thus our project is some part exploratory and charting an emerging field. “Over the last few years Data Science has really changed our concept of technology. Our lives are a lot easier as compared to ten years ago, and this is all because of data science. Data Science has really pulled the ends between fiction and technology.”, said Shweta Doshi, the co-founder of GreyAtom [5]. Our project helps children learn data science and fills the blank of data education in the children stage, as mentioned above.

Visualisation literacy is a must-have skill for people working on data science. Some teams have already carried out data visualisation literacy education in elementary school and kindergarten. Basak Alper et al. specifically contribute “C’est la vis”, an online platform for teachers and students to respectively teach and learn about pictographs and bar charts, and report on their initial observations of its use in grades K and 2 [6]. Besides, as an essential aid in data analysis, programming skills are crucial to data science practitioners. In [7], Neil Smith et al. founded “Code Club” in 2012, a network of after-school clubs in primary schools, aiming to teach children the basics of programming and fundamental CS concepts. If our teaching material gets good feedback and has practical use, then children can use the programming ability and visualisation literacy to independently work on data science. Imagine if the children of a country have mastered data science as an essential skill at their early education stage as equally well as reading and mathematics, how much progress will this country make in terms of its informationization level?

## 1.5 Feasibility

Our plan is divided into five parts, namely, making data sets, determining visualization methods, developing teaching materials, evaluating teaching materials, and optimizing teaching materials. Since our project is exploratory qualitative work, the evaluation process in the above plan is particularly important. For this we have set up two evaluation sessions. We start from the data set, visualization method, and teaching method suitable for children, and evaluate the teaching material twice (one of which is based on students’ feedback). Our plan is built around the target group (children) all the time, so it has a high feasibility.

## 1.6 Beneficiaries

In children’s education, the quality of the students’ teaching materials is crucial. The quality of the teaching material determines the expected outcome of this course to a certain extent. Hence, if our teaching material gets good feedback, it can help the future data education for children.

A good teaching material can guarantee a more satisfying data education for children. Therefore, we shall proceed from the children’s point of view (rather than in an adult’s intuitive for this subjective matter) when we implement every aspect of the plan, to ensure that our textbooks are useful and effective for data education for children.

## 2 Background

The general subject area of our project is children’s informatics education. In this field, many children-oriented courses are developing at a high speed.

[7] introduces Code Club, a network of after-school programming clubs offering on-campus programming education for primary school students within UK, run by technically-competent volunteers in conjunction with (generally technically-unskilled) teachers.

Online teaching material about computer science for children is available at [thestemlaboratory.com](http://thestemlaboratory.com). These teaching materials explain complicated computer science in ways that children can understand.

Some teams have already carried out data visualisation literacy education in primary school and kindergarten. Basak Alper et al. specifically contribute “C’est la vis”, an online platform for teachers and students to respectively teach and learn about pictographs and bar charts [6].

However, no organization is developing data education for children. The development of the field of data science has generated much discussion, enthusiasm, and investment within Colleges and Universities in recent years. Like any new discipline, Data Science brings new perspectives and new tools to address new questions requiring new perspectives on traditionally established concepts [1]. More and more people are paying attention to data science education and plan to conduct data science education in elementary schools.

## 3 Programme and Methodology

As presented above, we mentioned that our project is divided into five parts, i.e., making data sets, determining visualization methods, developing teaching materials, evaluating teaching materials, and optimizing teaching materials. The specific implementation steps of each plan are as follows:

- **Making the data set.** We first analyze what children might be interested in (e.g., pets). Then we search for the datasets associated with these regards and conduct preprocessing of the dataset.
- **Determining visualization methods.** We analyze the teaching methods suitable for children aged 10-12. Then develop appropriate data visualization methods from the determined teaching methods.
- **Developing teaching materials.** We combine the data set with data visualization methods mentioned above to develop the teaching material. We also briefly introduce “data for good” [8] in our teaching material.
- **Evaluating teaching materials.** We will evaluate the teaching material twice. The first evaluation is based on self - rates. Then we will organize an experimental class to do the second evaluation. We will refer to Richard and Rebecca’s method [9]. Taking the children’s cultural level into account, our questionnaire will appear in the form of multiple-choice questions. For open questions that can be difficult for participants to respond (for example, which part of the lecture do you think is difficult to understand?), we use audio recordings to document the responses, instead of the form of questionnaire.
- **Optimizing teaching materials.** After completing the evaluation of the teaching material, we will make targeted optimizations for the problems that arise during the evaluation.

We will focus on the characteristics of the children when implementing the plan. There is currently no authoritative teaching material suitable for children to learn data science. We will begin with the characteristics of children and explore the educational methods suitable for them, to develop practical teaching materials. These explorations show our contribution and the novelty of our project.

We set up four milestones in the process of completing the project, i.e., a) the completion of the first draft of the teaching material; b) the first optimization of the teaching material; c) the second optimization of the teaching material; d) the completion of the dissertation.

We have the following work packages: a) making the appropriate data set; b) exploring the appropriate teaching methods; c) exploring the appropriate visualisation methods; d) completing the first draft of the teaching material that can be used; e) completing the final teaching material based on the evaluation results; f) completing the dissertation.

Currently, there are some limits in our project planning. Due to the limitations of the people participating in the project and the duration of the project, it is not feasible for us to recurrently evaluate the teaching materials over two times. We will ask the professionals who is engaged in children's education to conduct a third evaluation if we get the chance during the implementation of the project.

## 4 Evaluation

We will evaluate the teaching material twice. The first evaluation is based on self - rates. We will conduct the first evaluation from the integrity of the teaching material and the accuracy of the expression.

The second evaluation will be based on the experimental classes. Our project is an exploratory qualitative work. The evaluation of the teaching material is mainly based on the feedback from the users (children) of the teaching material. We will refer to Richard and Rebecca's method [9]. The procedure of Richard and Rebecca's evaluation method is shown in figure 1. Taking the children's cultural level into account, our questionnaire will appear in the form of multiple-choice questions. For questions that are difficult to describe (for example, which part of the lecture do you think is difficult to understand?), we use audio recordings to record the answers, not in the form of a questionnaire.

## 5 Expected Outcomes

After two evaluations and optimization, we get the final edition of the teaching material. We hope that our teaching material is interesting, and children can learn most of the knowledge contained in the teaching material. The teaching material is expected to get good feedback and have practical applying. Combining the programming skills, visualization literacy, and data science they learned, children can perform data analysis independently, after learning the knowledge of our teaching material. In this way, our project can improve the situation that there is no widely acclaimed teaching material for children about data science.

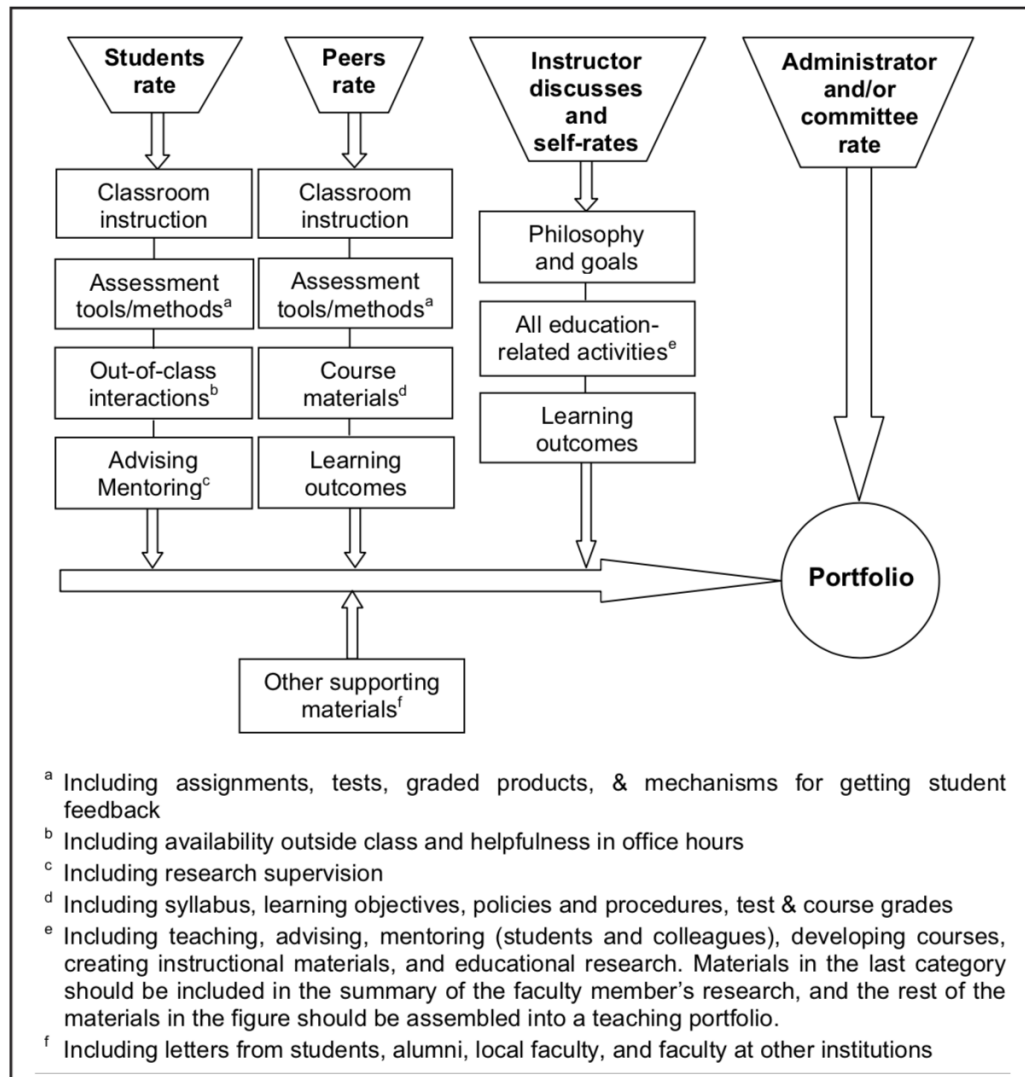


Figure 1: Richard and Rebecca's teaching performance evaluation model. Reference: [9]

## 6 Research Plan, Milestones and Deliverables

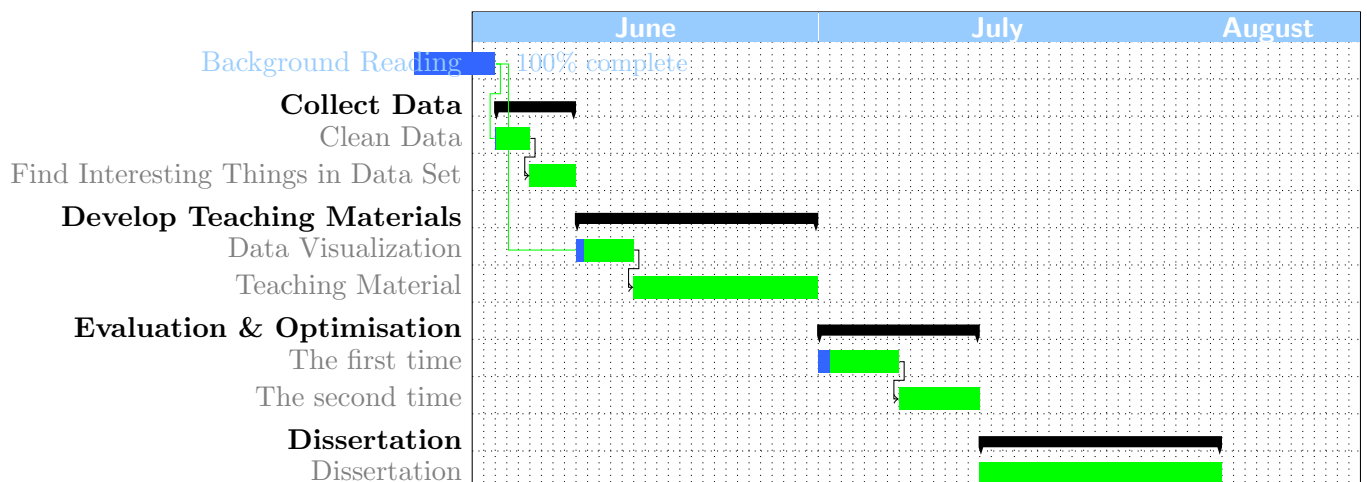


Figure 2: Gantt Chart of the activities defined for this project.

Milestone	Week	Description
$M_1$	4	The completion of the first draft
$M_2$	5	The first optimization of the teaching material
$M_3$	7	The second optimization of the teaching material
$M_4$	10	The completion of the dissertation

Table 1: Milestones defined in this project.

Deliverable	Week	Description
$D_1$	2	The appropriate data set
$D_2$	3	The appropriate teaching methods
$D_3$	4	The appropriate visualisation methods
$D_2$	5	The first draft of the teaching material
$D_3$	7	The final teaching material
$D_3$	10	The dissertation

Table 2: List of deliverables defined in this project.

## References

- [1] Jason Dedrick, Kenneth L Kraemer, and Eric Shih. Information technology and productivity in developed and developing countries. *Journal of Management Information Systems*, 30(1):97–122, 2013.
- [2] Dragomir. *Big Data and Its Impact on International Development*. [Online]. Available from: <https://wpmu.mah.se/nmict172group8/2017/09/25/big-data-bang-impact-international-development/>. [Accessed 25th SEPTEMBER 2017].
- [3] Stephanie C Hicks and Rafael A Irizarry. A guide to teaching data science. *The American Statistician*, 72(4):382–391, 2018.
- [4] Scottish Government. *Science, Technology, Engineering and Mathematics: education and training strategy*. [Online]. Available from: <https://www.gov.scot/publications/science-technology-engineering-mathematics-education-training-strategy-scotland/>. [Accessed 26th OCTOBER 2017].
- [5] Shweta Doshi. *Why is data science important?*. [Online]. Available from: <https://www.quora.com/Why-is-data-science-important>. [Accessed 26th JULY 2017].
- [6] Basak Alper, Nathalie Henry Riche, Fanny Chevalier, Jeremy Boy, and Metin Sezgin. Visualization literacy at elementary school. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 5485–5497. ACM, 2017.
- [7] Neil Smith, Clare Sutcliffe, and Linda Sandvik. Code club: bringing programming to uk primary schools through scratch. In *Proceedings of the 45th ACM technical symposium on Computer science education*, pages 517–522. ACM, 2014.
- [8] Jeannette M. Wing. Data for good: Abstract. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, KDD ’18, pages 4–4, New York, NY, USA, 2018. ACM.
- [9] Richard M Felder and Rebecca Brent. How to evaluate teaching. *Chemical Engineering Education*, 38(3):200–215, 2004.